An overview of a nonparametric estimation method for Lévy processes

José Enrique Figueroa-López

Purdue University

Joint work with Christian Houdré

Lévy Processes in Finance and Econometrics.

October 19th, 2005

Outline

I. Motivation: Asset price modeling by exponential Lévy processes

II. Nonparametric estimation of Lévy densities based on continuous observations

III. Implementation based on discrete observations

IV. Numerical illustration

V. Long-run asymptotics and Robustness: Oracle inequality and minimax performance

VI. Open questions

I. Motivation

1. Exponential Lévy models are some of the simplest and most practical alternatives to the shortfalls of the Black-Scholes model.

2. Exponential Lévy model can capture some stylized empirical features of historical returns; e.g. heavy tails, high kurtosis, and asymmetry.

3. Limitations: Lack of stochastic volatility, leverage, long-memory of absolute values, etc.

4. As a good “first-order” approximation model: Ideal to test statistical calibration methods.

5. Statistical issues: computationally expensive and numerically instable to estimate by traditional likelihood-based methods

II. Nonparametric estimation based on continuous observations

1. Why to study estimation based on continuous observations?
   (a) Provides Benchmarks for estimation methods based on discrete observations.
   (b) Serve as devises to construct discrete-based procedures by approximating the statistics underlying the continuous-based methods.
2. Standing assumptions:
   (A) Statistics of the form

$$\sum_{t \leq T} \varphi \left(\Delta X_t \right),$$

can be computed for $\varphi : \mathbb{R} \backslash \{0\} \to \mathbb{R}$.

(B) On an *estimation window* $D \subset \mathbb{R} \backslash \{0\}$, the Lévy measure $\nu(dx)$ can be written as

$$\nu(dx) = s(x)\, \eta(dx), \quad x \in D,$$

for a known measure η such that s is positive, bounded, and

$$\int_D s^2(x) \eta(dx) < \infty.$$

Notation: s *is said to be the Lévy density on* D *of the process with respect to the reference measure* η.

3. Formulation of the problem:

   Estimate *directly* the Lévy density s by non-parametric methods; that is, we make only qualitative assumptions about s.

4. Examples of models satisfying the assumption:

   (a) Standard continuous Lévy densities: $\nu(dx) = s(x)\, dx$.

   Estimate away from the origin (e.g. $D = (a, \infty)$ with $a > 0$):

   (b) Tempered stable processes:

   $\nu(dx) = |x|^{-\alpha-1} q(x) dx$, with $0 < \alpha < 2$, and q continuous and bounded.

   Estimate $s(x) = |x|^{6-\alpha-1} q(x)$ around the origin (e.g. $D = (-a, a)$) with respect to $\eta(dx) = x^{-6} dx$.

5. Main ideas:

(a) Approximation by finite-dimensional linear models:

$$s(x) \approx \beta_1 \varphi_1(x) + \cdots + \beta_n \varphi_n,$$

where the φ's are known functions. The space

$$\mathcal{S} := \{\beta_1 \varphi_1(x) + \cdots + \beta_n \varphi_n : \quad \beta_1, \ldots, \beta_n \quad \text{reals}\}$$

is called an (approximating) linear model.

(b) Two problems to solve:

i. Estimate a good element of $\mathcal{S}$. [Projection estimation]

ii. Determine a good approximating model from a collection of linear models, which are dense in general classes of functions; e.g. Splines and Wavelets. [Model selection]

6. The general method of estimation:

(a) Simplifying assumption: $\int \varphi_i^2(x)\eta(dx) < \infty$.

$\varphi_1, \ldots, \varphi_n$ are assumed to be orthonormal on D wrt η:

$$\int_D \varphi_i(x)\varphi_j(x)\eta(dx) = \delta_{i,j}.$$

(b) Projection estimator on $\mathcal{S}$:

$$\hat{s}(x) := \hat{\beta}(\varphi_1)\varphi_1(x) + \cdots + \hat{\beta}(\varphi_n)\varphi_n(x),$$

where

$$\hat{\beta}(\varphi) := \frac{1}{T}\sum_{t\leq T}\varphi(\Delta X_t).$$

(c) Basic Properties

i. Unbiasedness:

$\hat{s}$ is an unbiased estimator of the orthogonal projection of s on $\mathcal{S}$:

$$s^{\perp} = \beta(\varphi_1)\varphi_1(x) + \cdots + \beta(\varphi_n)\varphi_n(x),$$

where

$$\beta(\varphi) := \int s(x)\varphi(x)\eta(dx).$$

ii. The integrated-mean square error:

$$\mathbb{E}\|s^{\perp} - \hat{s}\|^2 = \frac{1}{T}\sum_i \int_D \varphi_i^2(x)s(x)dx \stackrel{T\to\infty}{\longrightarrow} 0,$$

where $\|f\|^2 := \int_D f^2(x)\eta(dx)$.

iii. The risk of estimation:

$$\mathbb{E}\left[\|s-\hat{s}\|^2\right] = \|s-s^{\perp}\|^2 + \mathbb{E}\|s^{\perp}-\hat{s}\|^2,$$

(d) Data-driven Model Selection:

i. Intuition:

In principle, a “nice” Lévy density s can be approximated closely by general linear models such as splines or wavelet. However, $\mathcal{S}'$ larger than $\mathcal{S}$ imply smaller approximating error $\|s-s^{\perp}\|^2$ and larger variance $\mathbb{E}\|s^{\perp}-\hat{s}\|^2$.

ii. Objective: Accomplish a good trade off between the two terms.

iii. A sensible solution:

$$\mathbb{E}\left[\|s-\hat{s}\|^2\right] = \|s\|^2 + \mathbb{E}\left[-\|\hat{s}\|^2 + \mathrm{pen}\left(\mathcal{S}\right)\right]$$

where

$$\mathrm{pen}\left(\mathcal{S}\right) := \frac{2}{T^2}\sum_{t\leq T}\bar{\varphi}^2_{\mathcal{S}}\left(\Delta X_t\right), \quad \text{and} \quad \bar{\varphi}^2_{\mathcal{S}}(x) := \sum_i \varphi_i^2(x).$$

Key observation: $-\|\hat{s}\|^2 + \mathrm{pen}\left(\mathcal{S}\right)$ is Observable !!

iv. The penalized projection estimator, on a collection of linear models $\mathcal{M}$ corresponding to a penalty function $\mathrm{pen} : \mathcal{M} \to [0,\infty)$, is the projection estimator $\tilde{s}$ that attains the minimum of

$$\min_{\mathcal{S}\in\mathcal{M}}\left\{-\|\hat{s}_{\mathcal{S}}\|^2 + \mathrm{pen}(\mathcal{S})\right\},$$

where $\hat{s}_{\mathcal{S}}$ is the projection estimator on $\mathcal{S}$.

III. Implementation based on discrete observations

1. Problem:

   Estimate

$$\hat{\beta}\,(\varphi) := \frac{1}{T} \sum_{t \leq T} \varphi(\Delta X_t),$$

   based on n equally spaced observations on $[0, T]$:

   $X_{T/n}, X_{2T/n}, \ldots, X_T$.

2. A simple solution:

$$\hat{\beta}_n\,(\varphi) := \frac{1}{T} \sum_{k=1}^{n} \varphi\left(X_{t_k} - X_{t_{k-1}}\right),$$

   where $t_k = T\frac{k}{n}$, for $k = 1, \ldots, n$.

3. Elementary Properties:

Suppose that $D := [a, b] \subset \mathbb{R}\backslash\{0\}$ and φ is a piece-wise continuous on D. Then,

(a) $\hat{\beta}_n(\varphi) \quad \xrightarrow{\mathfrak{D}} \quad \hat{\beta}(\varphi)$, as $n \to \infty$

(b) $\lim_{n\to\infty} \mathbb{E}\left[\hat{\beta}_n(\varphi)\right] = \beta(\varphi) := \int \varphi(x)s(x)\eta(dx)$

(c) $\lim_{n\to\infty} \mathrm{Var}\left[\hat{\beta}_n(\varphi)\right] = \frac{1}{T}\beta\left(\varphi^2\right)$.

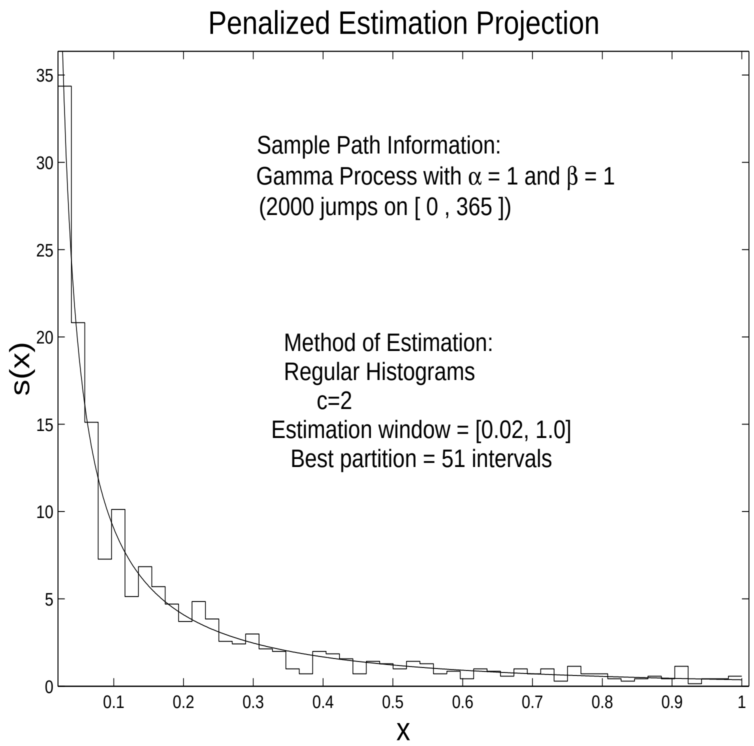
(d) If $\hat{s}^{(n)}$ is the approximate projection estimator of s on a given linear model $\mathcal{S}$, based on n discrete regular observations on $[0, T]$,

$$\lim_{n\to\infty} \mathbb{E}\left[\|\hat{s}^{(n)} - s\|^2\right] = \mathbb{E}\left[\|\hat{s} - s\|^2\right].$$

IV. A numerical example

- Gamma Lévy process with Lévy density $p(x) = \frac{\alpha}{x} e^{-x/\beta}$,
- Histogram type estimators (i.e. the φ's are indicator functions)

1. Model selection method based on a finite sample of jumps



2. Model selection method based on equally spaced observations

Δt	NPE-LSF		MLE	
1	1.01	1.46	.997	.995
.5	1.03	1.09	.972	.978
.1	.944	.995	1.179	.837
.01	.969	.924	6.92	.5

Table 1: Terminology: Δt is the time span between sampling observations.

NPE-LSF= "Least-square errors fit of the model $\frac{\alpha}{x}e^{-x/\beta}$ to the nonparametric histogram estimator".

Simulation: Results based on the simulations of 36500 jumps on $[0, 365]$

V. Theoretical Qualities.

1. Risk bound:

   (a) $\mathcal{M}$ - collection of linear models.

   (b) $D_{\mathcal{S}} := \| \sum_i \varphi_i^2(\cdot) \|_\infty$, where $\{\varphi_i\}_i$ is an orthonormal basis for $\mathcal{S}$.

   (c) $\mathcal{M}_T := \{\mathcal{S} \in \mathcal{M} : D_{\mathcal{S}} \leq T\}$.

   The penalized projections estimator $\widetilde{s}_T$ on $\mathcal{M}_T$ satisfies

$$\mathbb{E}\left[\|s - \tilde{s}_{T}\|^2\right]$$

$$\leq C \inf_{m \in \mathcal{M}_T} \left\{\|s - s^{\perp}_{\mathcal{S}_m}\|^2 + \mathbb{E}[\text{pen}(m)]\right\} + \frac{C'}{T},$$

   for a constant C (independent of $\mathcal{M}$, s and T), and C' (independent $\mathcal{M}$ and T.).

2. Oracle Inequality: The risk of $\tilde{s}_T$ is comparable (up to a constant) to the best possible risk using projection estimators:

$$\mathbb{E}\left[\|s - \tilde{s}_T\|^2\right] \leq C \inf_{\mathcal{S}\in\mathcal{M}_T} \mathbb{E}\left[\|s - \hat{s}_{\mathcal{S}}\|^2\right] + \frac{C'}{T}$$

3. Long-run rate of convergence: Let $\mathcal{B}$ be the Besov Space $\mathcal{B}^{\alpha}_{\infty}(\mathbb{L}^{\infty}[a,b])$, for $[a,b] \subset \mathbb{R}\backslash\{0\}$. If $s \in \mathcal{B}$ and $\widetilde{s}_T$ is penalized projection estimator on the collection of piece-wise polynomials of degree at most $\lfloor \alpha \rfloor$ on regular partitions of $[a,b]$, then

$$\limsup_{T\to\infty} T^{2\alpha/(2\alpha+1)} \sup_{s\in\Theta} \mathbb{E}\left[\|s - \widetilde{s}_T\|^2\right] < \infty,$$

where $\Theta := \left\{s \in \mathcal{B} : \|s\|_{\infty} < R, \quad \text{and} \quad |s|_{\mathcal{B}^{\alpha}_{\infty}} < L\right\}$.

4. Comparison to minimax risks:

Let $[a,b]$ be a closed interval of $\mathbb{R}\backslash\{0\}$, then

$$\liminf_{T\to\infty} T^{2\alpha/(2\alpha+1)} \left\{ \inf_{\hat{s}_T} \sup_{s\in\Theta} \mathbb{E}\left[\|\hat{s}_T - s\|^2\right] \right\} > 0, \qquad (1)$$

where the infimum is over all estimators $\hat{s}_T$ based on the jumps of the Lévy process $\{X(t)\}_{0\le t\le T}$ with sizes falling on $[a,b]$.

Summary

We developed estimation and model selection schemes for the Lévy density of a Lévy process:

- Flexible: it can be used histograms, splines, wavelets, etc.
- Model free
- Easily implementable
- Reliable and robust: Oracle inequality and adaptivity; i.e. asymptotically comparable to minimax estimators on classes of smooth Lévy densities
- Applicable to estimate and assess specific parametric models via a least-squares fit method

VI. Some open problems

(1) Investigate asymptotics of non-parametric discrete-data based estimators which are uniform on classes of Lévy densities, as well as their comparisons to the long-run minimax risk. These asymptotics will be as both the frequency of the observations and the time horizon increase.

(2) Extend the minimax and rate of convergence results when considering arbitrarily small jumps and when estimating around the origin, where the Lévy density usually blows up. How to assess the quality of estimation around the origin? How to approximate functions that blows up?

(3) Apply similar ideas to more realistic models such as time-changed Lévy processes and jump-diffusion models with stochastic volatility. For instance, estimation of the functional parameters driving both the random clock and the jump process.